



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/649,909

08/26/2003

Satyanarayana Dharanipragada

N0484.70762US00

5755

23628 7590 09/21/2009
WOLF GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

EXAMINER

COLUCCI, MICHAEL C

ART UNIT

PAPER NUMBER

2626

MAIL DATE

DELIVERY MODE

09/21/2009

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No. 10/649,909	Applicant(s) DHARANIPRAGADA ET AL.	
	Examiner MICHAEL C. COLUCCI	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 20 July 2009.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-27 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-27 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Response to Arguments

1. Applicants arguments with respect to claims 1-27 have been considered but are moot in view of the new grounds of rejection. Examiner withdrawn Chang and has incorporated Neti et al. US 5953701 A (hereinafter Neti) as the primary reference to address amendments within claim(s) 1-27. See rejection.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

3. Claims 1-27 rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. "*A speech recognition model of phoneme models*" and "phoneme model pairs" are not necessarily well known in the art nor appear to be found in the specification. Regardless, Examiner has found art to address these limitations, wherein Examiner believes that a "*pair*" appears inherent, given the context of the present invention, whereby two pieces of data are required for comparison. Re

Art Unit: 2626

"A speech recognition model of phoneme models", while lacking support, Examiner construes this as a speech recognition phoneme model intrinsically containing phonemes (i.e. phoneme model used in speech recognition).

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 1, 2, 5, 6, 7, 10-12, 15, and 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Yang US 20010010039 A1 (hereinafter Yang).

Re claims 1, 6, 11, and 16, Neti teaches a method for generating speech recognition models, the method comprising:

receiving a female speech recognition model of phoneme models based on the female set of recorded phonemes training data (Col. 5 lines 9-21, Fig. 4);

receiving a male speech recognition model of phoneme models based on the male set of recorded phonemes training data (Col. 5 lines 9-21, Fig. 4);

Art Unit: 2626

determining a difference in model information between pairs of corresponding phoneme models of the female speech recognition model and the male speech recognition model Col. 5 lines 9-21);

creating a gender-independent speech recognition model that includes a gender-independent phoneme model based on if a pair of corresponding phoneme models of the female speech recognition model and the male speech recognition model (Col. 5 lines 9-21) when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant (Col. 1 lines 33-47)

However, Neti fails to teach phoneme training data and phoneme models

Yang teaches a Mandarin Chinese speech recognition apparatus comprises, a speech signal filter for receiving a speech signal and creating a filtered analogue signal, an analogue-to-digital (A/D) converter connected to the speech signal to a digital speech signal, a computer connected to the A/D converter for receiving and processing the digital signal, a pitch frequency detector connected to the computer for detecting characteristics of the pitch frequency of the speech signal thereby recognizing tone in the speech signal, a speech signal pre-processor connected to the computer for detecting the endpoints of syllables of speech signals thereby defining a beginning and ending of a syllable, and a training portion connected to the computer for training an initial part PSV model and a final part PSV model and for training a syllable model based on trained parameters of the initial part PSV model and the final part PSV model (Yang [0016]).

Art Unit: 2626

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate phoneme training data and phoneme models as taught by Yang to allow for defining a beginning and ending of a syllable, wherein characteristics such as pitch and tone are used to find differences between phonemes (Yang [0016]) in both male and female voices.

Re claims 2, 7, and 12, Neti teaches the method at least one computer readable medium of claim 1, further comprising removing each of the phoneme models of the pair of corresponding phoneme models from the female speech recognition model and the male speech recognition model (Col. 5 lines 9-21, Fig. 4 & 5, processor 44 outputs recognized speech based on female dependent models, male dependent models, and male and female independent models 46 ad 48) when the difference in model information between the phoneme models is insignificant (Col. 1 lines 33-47).

Note: Examiner finds support for the act of “*removing*” such as “*the processor 108 removes the separate female models 110 and male models 112 that are determined to have insignificant differences from one another. The final result from the processor 108 contains female models 110 derived from female training data 104, male models 112 derived from male training data 106, and gender independent models 114 derived from both the female and male training data 104 and 5 106, wherein the female models 110 and male models 112 are significantly different from each other*” (present invention spec. page 3 line 30 - page 4 line 6).

However, Neti fails to teach phoneme training data and phoneme models

Yang teaches a Mandarin Chinese speech recognition apparatus comprises, a speech signal filter for receiving a speech signal and creating a filtered analogue signal, an analogue-to-digital (A/D) converter connected to the speech signal to a digital speech signal, a computer connected to the A/D converter for receiving and processing the digital signal, a pitch frequency detector connected to the computer for detecting characteristics of the pitch frequency of the speech signal thereby recognizing tone in the speech signal, a speech signal pre-processor connected to the computer for detecting the endpoints of syllables of speech signals thereby defining a beginning and ending of a syllable, and a training portion connected to the computer for training an initial part PSV model and a final part PSV model and for training a syllable model based on trained parameters of the initial part PSV model and the final part PSV model (Yang [0016]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate phoneme training data and phoneme models as taught by Yang to allow for defining a beginning and ending of a syllable, wherein characteristics such as pitch and tone are used to find differences between phonemes (Yang [0016]) in both male and female voices.

Re claims 5, 10, and 15, Neti teaches method of claim 1, wherein the female speech recognition model, male speech recognition model, and gender-independent speech recognition model are Gaussian mixture models (Neti Col. 3 lines 50-67).

6. Claims 3, 4, 8, 9, 13, and 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Yang US 20010010039 A1 (hereinafter Yang) and further in view of Kanevsky et al. US 6529902 (hereinafter Kanevsky).

Re claims 3, 8, and 13, Neti in view of Yang fails to teach the method of claim 1, wherein determining the difference in model information includes calculating a Kullback Leibler distance between the first speech recognition model and second speech recognition model.

Kanevsky et al. teaches that for two different sets, one can define a Kullback-Leibler distance using the frequencies of the sets. [With the distance] one can check which pairs of topics are sufficiently separated from each other. Topics that are close in this metric could be combined together (Kanevsky Col. 12, lines 42-47).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Yang to incorporate the determining the difference in model information includes calculating a Kullback Leibler distance between the first speech recognition model and second speech recognition model as taught by Kanevsky to allow for an improved language modeling for off-line automatic speech decoding and machine translation (Kanevsky Col. 2, lines 50-52).

Re claims 4, 9, and 14, Neti in view of Yang fails to teach the method of claim 3, wherein whether the model information is insignificant is based on a threshold Kullback Leibler distance quantity.

Kanevsky teaches the Kullback-Leibler distance (Kanevsky Col. 5, lines 9-11) between any two topics is at least h , where h is some sufficiently large threshold, also they teach (Kanevsky Col. 12, lines 44-47) that while using the Kullback-Leibler distance, one can check which pairs of topics are sufficiently separated from each other, and that topics that are close in this metric could be combined together).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Yang to incorporate whether the model information is insignificant is based on a threshold Kullback Leibler distance quantity as taught by Kanevsky to allow for an improved language modeling for off-line automatic speech decoding and machine translation, wherein a sufficiently large threshold indicates separate or combinational probabilities (Kanevsky Col. 2, lines 50-52).

7. Claims 17-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Wark US 20030231775 (hereinafter Wark) and further in view of Yang US 20010010039 A1 (hereinafter Yang).

Art Unit: 2626

Re claims 17, 21, and 24, Neti teaches Wark teaches a system for recognizing speech data from an audio stream originating from one of a plurality of data classes ([0094]) system comprising:

a computer processor (Col. 6 lines 24-49);

a receiving module configured to receive a current feature vector of the audio stream (Col. 6 lines 24-49);

a first computing module configured to compute a current vector probability (Col. 3 lines 50-67) that the current feature vector belongs to one of the plurality of data classes (Col. 5 lines 9-21);

wherein the plurality of data classes include a first speech recognition model based on recorded phonemes originating from a first set of speakers, a second speech recognition model based on recorded phonemes from a second set of speakers, and a third speech recognition model that includes phoneme models based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers having insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes (Col. 5 lines 9-21, Fig. 4 & 5), each of the first speech recognition model and the second speech recognition model lacking the phoneme models of the third speech recognition model based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers having insignificant differences in model information between the recorded phonemes of the pairs of corresponding recorded phonemes (Col. 1 lines 33-47 & Fig. 4).

However, Neti fails to teach a second computing module configured to compute an accumulated confidence level that the audio stream belongs to one of the plurality of data classes based on the current vector probability and on previous vector probabilities;

a weighing module configured to weigh class models based on the accumulated confidence; and

a recognizing module configured to recognize the current feature vector (based on the weighted class models; and

Wark teaches classification of homogeneous segments, a number of statistical features are extracted from each segment. Whilst previous systems extract from each segment a feature vector, and then classify the segments based on the distribution of the feature vectors, method 200 divides each homogenous segment into a number of smaller sub-segments, or clips hereinafter, with each clip large enough to extract a meaningful feature vector f from the clip. The clip feature vectors f are then used to classify the segment from which it is extracted based on the characteristics of the distribution of the clip feature vectors f . The key advantage of extracting a number of feature vectors f from a series of smaller clips rather than a single feature vector for a whole segment is that the characteristics of the distribution of the feature vectors f over the segment of interest may be examined. Thus, whilst the signal characteristics over the length of the segment are expected to be reasonably consistent, by virtue of the segmentation algorithm, some important characteristics in the distribution of the feature

Art Unit: 2626

vectors f over the segment of interest is significant for classification purposes (Wark [0094])

Further, Wark teaches the ability to decide whether the segment should be assigned the label of the class with the highest score, or labeled as "unknown", a confidence score is calculated. This is achieved by taking the difference of the top two model scores $s_{p,q}$ and $s_{q,p}$, and normalizing that difference by the distance measure D_{pq} between their class models p and q . This is based on the premise that an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model. With further apart models, the model scores $s_{p,q}$ should also be well separated before the segment is assigned the class label of the class with the highest score (Wark [0146] & Fig. 4, adjacent, previous and current segment/frame).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a second computing module configured to compute an accumulated confidence level that the audio stream belongs to one of the plurality of data classes based on the current vector probability and on previous vector probabilities, a weighing module configured to weigh class models based on the accumulated confidence and a recognizing module configured to recognize the current feature vector (based on the weighted class models as taught by Wark to allow for normalization of a difference by a distance, whereby an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model (Wark [0146])), wherein a confidence score or score is used to better classify speech, whereby segments of feature vectors are classified, making important

Art Unit: 2626

characteristics in adjacent, current, and previous frames in the distribution of the feature vectors more apparent (Wark [0094]).

However, Neti in view of Wark fails to teach phoneme training data and phoneme models

Yang teaches a Mandarin Chinese speech recognition apparatus comprises, a speech signal filter for receiving a speech signal and creating a filtered analogue signal, an analogue-to-digital (A/D) converter connected to the speech signal to a digital speech signal, a computer connected to the A/D converter for receiving and processing the digital signal, a pitch frequency detector connected to the computer for detecting characteristics of the pitch frequency of the speech signal thereby recognizing tone in the speech signal, a speech signal pre-processor connected to the computer for detecting the endpoints of syllables of speech signals thereby defining a beginning and ending of a syllable, and a training portion connected to the computer for training an initial part PSV model and a final part PSV model and for training a syllable model based on trained parameters of the initial part PSV model and the final part PSV model (Yang [0016]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Wark to incorporate phoneme training data and phoneme models as taught by Yang to allow for defining a beginning and ending of a syllable, wherein characteristics such as pitch and tone are

Art Unit: 2626

used to find differences between phonemes (Yang [0016]) in both male and female voices.

Re claims 18, 22, and 25, Neti teaches the method of claim 17, wherein computing the current vector probability includes estimating a posteriori class probability for the current feature vector (Col. 2 lines 1-8))

Re claims 19, 23, and 26, Neti fails to teach the method of claim 17, wherein computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities.

Wark teaches classification of homogeneous segments, a number of statistical features are extracted from each segment. Whilst previous systems extract from each segment a feature vector, and then classify the segments based on the distribution of the feature vectors, method 200 divides each homogenous segment into a number of smaller sub-segments, or clips hereinafter, with each clip large enough to extract a meaningful feature vector f from the clip. The clip feature vectors f are then used to classify the segment from which it is extracted based on the characteristics of the distribution of the clip feature vectors f . The key advantage of extracting a number of feature vectors f from a series of smaller clips rather than a single feature vector for a whole segment is that the characteristics of the distribution of the feature vectors f over the segment of interest may be examined. Thus, whilst the signal characteristics over

Art Unit: 2626

the length of the segment are expected to be reasonably consistent, by virtue of the segmentation algorithm, some important characteristics in the distribution of the feature vectors f over the segment of interest is significant for classification purposes (Wark [0094])

Further, Wark teaches the ability to decide whether the segment should be assigned the label of the class with the highest score, or labeled as "unknown", a confidence score is calculated. This is achieved by taking the difference of the top two model scores $s_{p,q}$ and $s_{q,p}$, and normalizing that difference by the distance measure D_{pq} between their class models p and q . This is based on the premise that an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model. With further apart models, the model scores $s_{c,p}$ should also be well separated before the segment is assigned the class label of the class with the highest score (Wark [0146] & Fig. 4, adjacent, previous and current segment/frame).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities as taught by Wark to allow for normalization of a difference by a distance, whereby an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model (Wark [0146]), wherein a confidence score or score is used to better classify speech, whereby segments of feature vectors are classified, making important characteristics in adjacent, current, and previous frames in the distribution of the feature vectors more apparent (Wark [0094]).

Re claims 20 and 27, Neti teaches the method of claim 17, further comprising determining if another feature vector is available for analysis (Col. 6 lines 24-49).

Conclusion

8. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-

Art Unit: 2626

270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner
AU 2626
(571)-270-1847
Examiner FAX: (571)-270-2847
Michael.Colucci@uspto.gov

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626